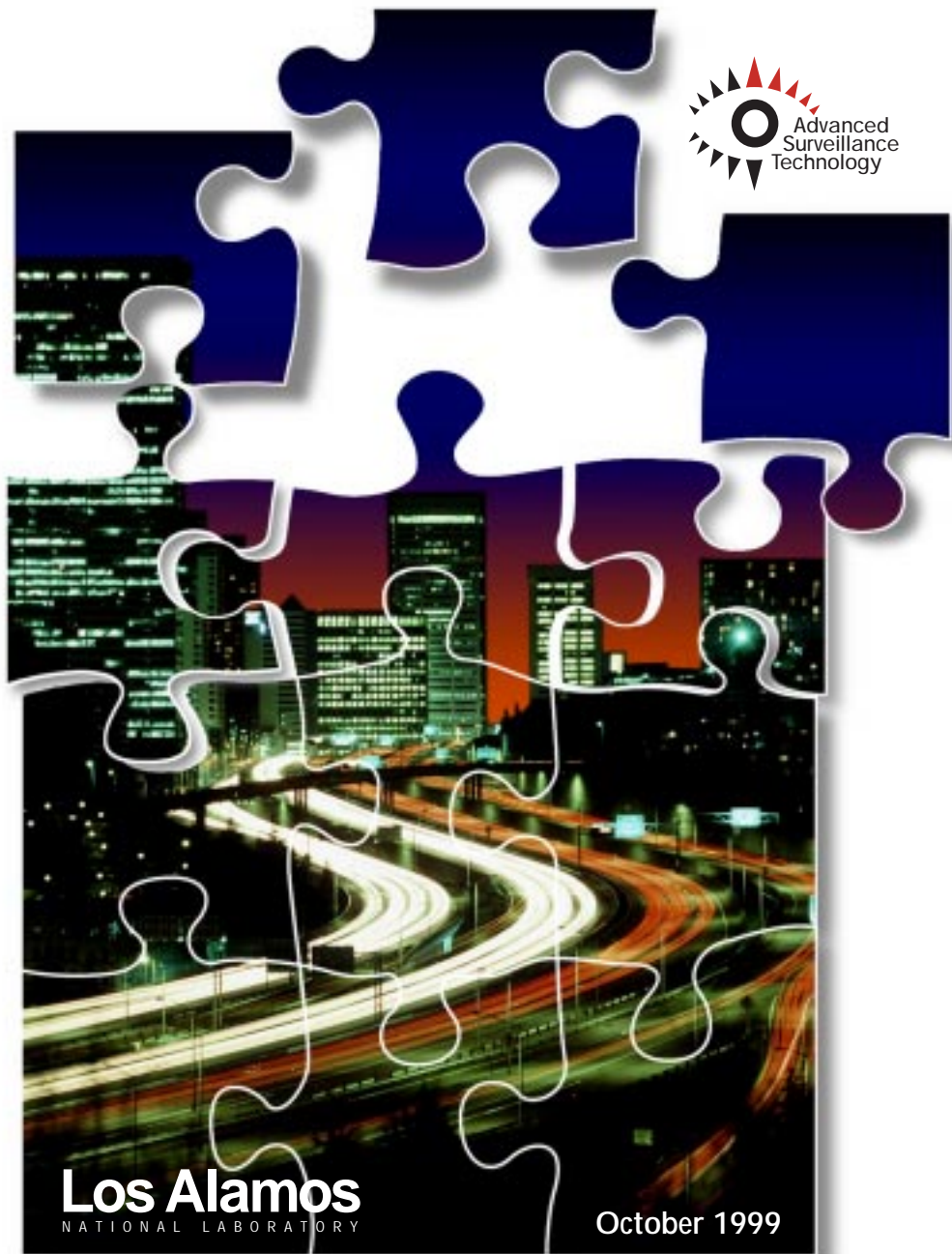


BITS

AMISS compiles more than you think.



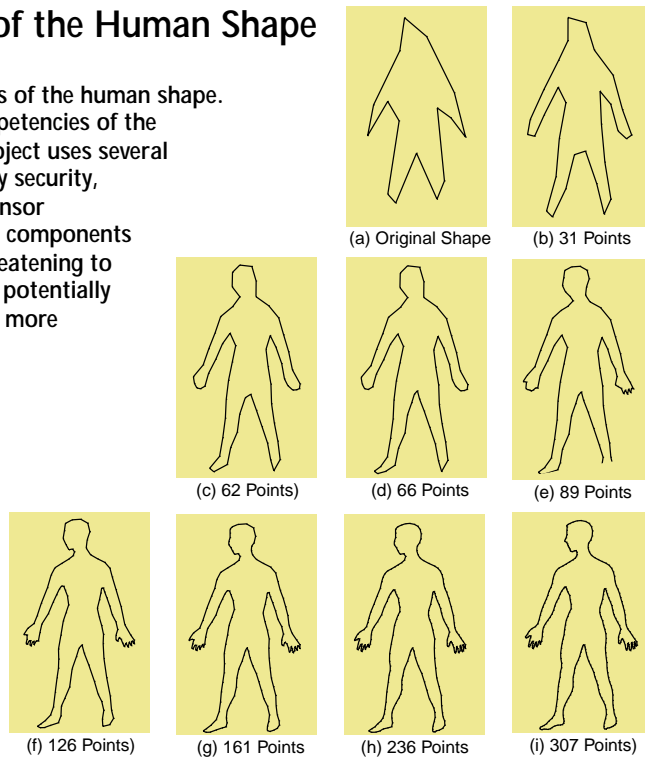
See Inside

Data Analysis in the Advanced Surveillance Technology Initiative: AMISS Testbed
Transaction Integrity and Validity in MC&A Databases

Successive Polygonal Approximations of the Human Shape

These images represent successive polygonal approximations of the human shape. Shape Recognition and Identification is one of the core competencies of the Advanced Surveillance Technology Initiative project. This project uses several different artificial intelligence technologies to increase facility security, safeguards, and safety. The testbed, called Adaptive Multisensor Integrated Security System (AMISS), integrates several R&D components that specifically targets actions that could be considered threatening to the facility and uses pattern recognition to distinguish these potentially threatening actions and behavior from normal behavior. For more information about this research and development, contact Sharon Seitz, Project Leader for the Advanced Surveillance Technology Initiative, at (505) 665-6812 or send e-mail to sharons@lanl.gov.

(Editor's note: This issue of BITS has two articles related to the Advanced Surveillance Technology Initiative. These articles describe collaborative research and development including data mining, data analysis, and developing expert systems to improve the accuracy of the information stored and reported about the nuclear materials at LANL. See these articles: *Data Analysis in the Advanced Surveillance Technology Initiative: AMISS Testbed* and *Transaction Integrity and Validity in MC&A Databases*.)



Produced by the Computing, Information, and Communications (CIC) Division

Design: Donald Montoya and Julie Medina (CIC-1)

Managing Editor: Denise Sessions, denise@lanl.gov (CIC-1)

Illustration: Dave Delano-Art/Illustration

Photography: John Flower (CIC-9)

Printing: Imaging Services Group (CIC-9)

BITS Contributors' Board: Hal Marshall, CIC-6; Kimberlyn Mousseau, CIC-15; Gina Fisk, CIC-5; Kathleen Jackson, CIC-3; Kei Davis, CIC-19; Nikki Gaedecke, CIC-6; David Van Etten, CIC-6, Denise Sessions, CIC-1/CIC-6, and Don Willerton, CIC-6 Group Leader.

Attention Subscribers: There are two ways you can update your subscription data. E-mail bitsupdate@lanl.gov, or fill out the form inside this issue. Internal LANL subscribers please note that until we have a mechanism in place to verify our subscription database with the LANL Employee Information System we will rely on you to update your mailstop. Thanks for keeping us updated on subscription data.

Table of Contents

Information Systems	2
MathML: A Kind of "TeX for the Web"	2
 Desktop Computing	 4
A Primer for Developing Online Help	4
News about Laboratory Electronic Mail and Policies	6
 Infrastructure	 9
Data Analysis in the Advanced Surveillance Technology Initiative: AMISS Testbed	9
Transaction Integrity and Validity in MC&A Databases	12
 High Performance Computing	 16
Frameworks Are Models, Too	16
Lecture Review: Automated OS Install HOWTO: Linux Cluster via Gigabit Ethernet	20
 What's Happening	 22
A New Look and Feel to xxx.lanl.gov	22
Research Library	23
Technical and Advanced Technical Computer Training Courses	24
BITS Subscription/Comment Form	25
1999 12-Month Index	27

MathML: A Kind of “TeX for the Web”

by Tad Lane, Information Architecture Standards Editor, CIC-1 Communication Arts and Services

In last month's BITS article on Web Content Architecture, I took a high-level view at how eXtensible Markup Language (XML) documents pull in other Web content elements, such as images, scripts, and style sheets. I discussed how XML is a language for writing other markup languages, and I briefly mentioned some examples.

For this month, as a further demonstration of how XML works, I'd like to follow up with a closer look at one of the XML languages, showing some example markup and example renderings. More specifically, I'll focus on Mathematical Markup Language (MathML), one of the XML languages that is being promoted by the World Wide Web Consortium (W3C).

MathML still requires special tools for rendering, but such tools are available and the language can be used to meet special needs today. The IntraLab99 Web Conference in May, for example, included a live demonstration of MathML using the W3C Amaya browser.

MathML

MathML is an XML language that can be used to describe both the rendering and the meaning of mathematical expressions. It can be thought of as a kind of “TeX for the Web”, enabling us to include complex math in our Web

pages. It currently requires hard-coded browser extensions or plug-ins to be correctly interpreted, though style sheets may eventually be able to deliver the rendering instructions.

As a simple example, we can take Einstein's $E = mc^2$ equation. If we want to describe that in terms of how it should be displayed on a computer, we can insert the following presentation markup into our familiar HTML markup:

```
<math>
  <mrow>
    <mi>E</mi>
    <mo>=</mo>
    <mi>m</mi>
    <msup>
      <mi>c</mi>
      <mn>2</mn>
    </msup>
  </mrow>
</math>
```

The first thing to note here is that every tag has a closing tag, meeting the XML requirement that every tag close. The markup then “reads” as follows:

- `$...$` mark the start and end of the mathematical expression.
- `<mrow>...</mrow>` mark the start and end of the row to be displayed.
- `<mi>e</mi>` is the mathematical identifier of “E”.

- `<mo>=</mo>` is the mathematical operator of “=”.
- `<mi>m</mi>` is the mathematical identifier of “m”.
- `<msup>...</msup>` mark the start and end of an item that will be displayed with a superscript.
- `<mi>c</mi>` is the mathematical identifier of “c”, which is what the superscripted item will be attached to.
- `<mn>2</mn>` is the mathematical number of “2”, which is the item that will be superscripted.

All of it describes the way the equation should be displayed.

If we would prefer, however, we could describe the same equation with the following content markup:

```
<math>
  <reln><eq/>
  <ci>E</ci>
  <apply><times/>
    <ci>m</ci>
    <apply><power/>
      <ci>c</ci>
      <cn>2</cn>
    </apply>
  </apply>
</reln>
</math>
```

This markup reads as follows:

- `$...$` again mark the beginning and end.
- `<reln>...</reln>` mark the beginning and end of a mathematical relationship.
- `<eq/>` signifies that the relationship is an equation. Per the XML specification, note how the `/>` closes the empty tag.
- `<ci>e</ci>` is the content identifier of "E", which is the first side of the equation.
- `<apply><times/>...</apply>` mark the beginning and end of a group of items that will be multiplied. With nesting, note that it is the second `</apply>`, the outside one, which closes this group. This group also constitutes the second side of the equation.
- `<ci>m</ci>` is the content identifier of "m", which is the first item that will be multiplied.
- `<apply><power/>...</apply>` mark the beginning and end of an exponentiation operation. The first item following will be put to the power of the second item. This also marks the group that will be multiplied by "m".
- `<ci>c</ci>` is the content identifier of "c", which is the item that will be exponentiated.
- `<cn>2</cn>` is the content number of "2", which is the power that "c" is raised to.

All of this describes the way the equation should be interpreted.

The default rendering would be the same for both the presentation markup and the content markup, but the interpretations are different. The presentation markup describes how to display the equation, and it is generally better for cases where we want to

control what the user sees. (It's also more generally supported at this time, which means it's sometimes our only real choice.) The content markup is better suited for cases where we want to avoid any ambiguity about meaning, and it opens the door to possibilities of presenting equations that can be interactively interpreted online (e.g., if I plug in an "m" value, I'll get the "E").

The above example, of course, could also be rendered using standard HTML ("E = mc²"), but MathML is not limited to such simple equations. Figure 1, for example, shows a screenshot of the first Lorentz transformation, taken from an Amaya browser interpreting MathML presentation markup.

$$t = \frac{t' + \frac{vx'}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Fig. 1. Screenshot of first Lorentz transformation.

For Further Information

The following BITS articles have addressed related subjects:

- Web Content Architecture: Taming the Tangle of Protocols (*September 1999*)
- eXtensible HTML: Preparing for the Next Phase of Web Markup (*August 1999*)
- Extending Web Documents: Getting Ready for XML (*March 1998*)

Further information about MathML, including links to MathML implementations, can be found at the W3C Web site at <http://www.w3.org/Math/>.

For additional information about the Information Architecture (IA) Web activities, please see our Web team page at <http://www.lanl.gov/projects/ia-lanl/area/web/>. For additional information about the IA Project in general, please see our project home page at <http://www.lanl.gov/projects/ia/>.

A Primer for Developing Online Help

by Sheila Molony, Technical Writer/
Editor, Software Documentation
Team, CIC-1 Communication Arts
& Services

Online Help is the content of instructions, reference materials, and other information that displays when you choose the **Help** button on a software application. Ideally, writers who work with the programming and design team during application development write the Help content. More often, however, the Help is nonexistent, or written by programmers who are often too close to the application to view it as a new user would, or written by a writer brought in at the end of the project.

Communication Arts and Services (CIC-1) has several writers who develop Online Help, user manuals, or other support materials for Laboratory-developed applications. This article describes some of the basic principles for developing Online Help.

The Evolution of Online Help

As the number and variety of retail software applications have grown over the last 10-15 years, writing user documents and support material has developed into a specialty. Specific software tools for organizing Online Help have been developed for writers, and informal standards for content, organization, and presentation of information have emerged. Because most applications were developed for Microsoft® operating systems, the *Microsoft Manual of Style for Technical*

Publications has become a prominent reference tool for Online Help writers. Online Help applications have primarily followed the "look" of the Microsoft model, which includes a Help window containing tabs for Contents, Index, and Search. But graphical interfaces and Web-based, cross-platform applications have increased the users' skill level and changed the way information is presented onscreen. New tools for writing Online Help for the Web are still being perfected, but the principles of organization of content still hold.

Types of Online Help

Research shows that users do not read Online Help unless they need help to perform a specific task. Therefore, an approach called "Task-Oriented" Help is primarily used by writers today. In Task-Oriented Help, the writer becomes familiar with the application and its users and then lists all the tasks that a user can perform on that application. Then the writer writes step-by-step instructions for accomplishing each task—what screens to move to, what buttons to choose, what fields to complete, etc. Each task is titled and listed in the table of contents so that a user can find it easily. Examples of Task-Oriented Help include "How to Print the XYZ Report" or "Plotting Table Data" or "Steps for Approving User-Submitted Forms."

The format for a task description has also become standard and includes a title, a brief introduction that contains useful information, an introduction to the steps, and then the steps themselves. An example of part of a task follows.

Task-Oriented Help Example: Plotting Table Data

Table numbers that you can plot are displayed as hyperlinks. Table hyperlinks are located in three places: the Tables column, the Title bar at the top of the Description frame, and the Table Summary list in the Description frame.

To plot table data:

1. On the EOS Main Page, locate the material of interest by name or Material Number. The tables available for that material are displayed in the Tables column.
2. Do one of the following:
 - In the Tables column, choose the table number hyperlink you want to plot. The Data Plot page will open showing the default plot and additional data.
 - Choose the Material Name or Sesame Number hyperlink to display the detail for that material in the Description frame. From the Tables list that appears, choose the table number hyperlink you want to plot.
3. Etc.

But task information alone is incomplete. If a user is following the steps to fill out a form but gets to a field or list of selections that he/she does not understand, another type of help is needed. It's called "Reference" Help.

Reference Help is simply a description of the individual elements of each page. The writer describes what each element is or does, even if the description is obvious. When describing the Name field, the description might be "Name of the person filling out this

form." Or, depending on the application, it might be "Your name." Every button, link, and field is described, and any other pertinent information included, such as the format for date entry or a character length limit.

Because we know users do not like to read much online (they scan quickly to find what they want), writers make the descriptions as concise as possible.

Reference Help Examples

Examples of Reference Help follow.

Add Itinerary Button

Allows you to add the new itinerary to your table of itineraries.

Are You Invited?

Indicates whether or not you are traveling as the result of an invitation. If invited, you must include the invitation as an attachment.

Business Involves

Indicates whether your business in the location involves sensitive topics, classified information, and/or international agreements.

Submit Button

Records the information and submits the form for approval.

Another type of Reference Help, known as "Context Sensitive" Help, automatically displays a pertinent pop-up message or window based on the location on the screen from which the user asks for help. Use of Context-Service Help is growing because it provides users with quick support where they need help and because the tools to develop it are becoming easier to use.

Task Help and Reference Help are two of the essential sections of Online Help content. In addition, a hyperlinked Help table of contents and contact for problem reporting are always included. Depending on the application, there are often other kinds of content, including an overview of features of the application, tables of information or definitions that relate specifically to the application, or hardware and software requirements for running the application. Usually, if you can see the Online Help, you've gotten to the application itself. But the requirements section often describes specific quirks or features that apply to different platforms. Also often included are an index, frequently asked questions, troubleshooting tips, and sections on how the application actually works.

With Online Help, writers can design Help systems to make finding information as easy as possible for the user. Including hyperlinks in the text or "See Also" or "Related Topics" hyperlinks sends the user to a new, related destination without having to return to the Help table of contents to find it. A user can go from a task instruction such as "In the Tables column, choose the table number hyperlink..." to a

Reference Help description of the page where the Tables column is located. The hyperlink could be the word "Tables" or a "See Also" called "EOS Main Page" at the bottom of the task instructions.

Understanding User's Needs

There's lots more to developing Online Help—working with the application developers to understand the typical user's needs and level of knowledge, playing with the application to understand how it works and define the tasks, and, ideally, testing the Online Help by having a few actual users provide feedback before development is final.

These are some of the principles CIC-1 software documenters apply to assure Online Help for Laboratory-developed applications is an effective tool to help users quickly become adept at new applications.

For more information about Online Help or for assistance in developing Help resources, contact Sheila Molony, CIC-1, at 5-1585.



News about Laboratory Electronic Mail and Policies

*by Emmy Hopson, CIC-5
Messaging Services*

Recent security events at the Laboratory have focused attention on electronic mail (e-mail) monitoring and on LANL policies regarding privacy and the control of electronic media. It seems like an appropriate time now to explain what is happening here and how it may affect us.

As all of us with LANL computers know, we now accept a policy stating that "users have no explicit or implicit expectation of privacy." The Laboratory has for some time used a Web scanning script to monitor visits to certain "inappropriate" Web sites. Now LANL is, also, somewhat randomly selecting a certain number of outgoing e-mail messages to be scanned for "classified" and, one assumes, "sensitive" information. With increased scrutiny by Congress and the Department of Energy, we can only expect that more tools will be developed to monitor our electronic behavior. For example, CIC-5 has been asked to use a product that blocks access to reported Web sites identified by an Internet service. Once software performance problems are alleviated, the service will probably be installed.

The Laboratory is preparing draft amendments to the Administrative Manual. These amendments will deal with privacy, waste, fraud, and abuse issues in regard to electronic communications in general, including e-mail.

E-Mail Policies

What are our policies? E-mail policies should be expected to be dynamic in nature and evolve along with the available technology and should address the following parameters:

- Personal use
- Employer monitoring
- Proper use
- Records retention

The following information has been pulled together from the Administrative Manual, the University of California Web pages, and the Internet Engineering Task Force (IETF) Web site.

Personal Use

Be careful about using electronic media for personal use. There is a draft policy pending UC approval to amend the Laboratory Administrative Manual to allow "occasional incidental personal use of electronic information resources, such as e-mail and the Internet," but personal use will be prohibited if any of the following result from this usage:

- Interference with the resources of the Laboratory.
- Burdening the Laboratory with additional costs.
- Interference with the user's employment or other obligations to the Laboratory.
- Unacceptable use.

Unacceptable use is defined as using the Laboratory resources for political activities, commercial purposes, financial gain, unlawful activities such as fraud, theft or gambling. In addition, viewing, reading, or distributing sexually explicit materials is prohibited. Unauthorized release of classified, privileged information, or restricted materials is prohibited.

Employer Monitoring

As stated above, outgoing e-mail and Web site access are being monitored. The Laboratory also has an emergency security policy that allows system managers to remove e-mail deemed as "classified" from accounts on the e-mail servers. This is done under the auspices of the Security (S) Division and is not read before deletion so that other systems are not contaminated.

Over 300 e-mail servers have been registered at the Laboratory as forwarding e-mail addresses. Many of these are single-user systems. System administrators on the "institutional" Lab systems managed by CIC-5 (esh-mail, cic-mail, nmt-mail, etc.) are occasionally required to view contents of mail file systems in order to maintain them. This may be done without authorization from the user but is accomplished with the least amount of perusal necessary to fix the problem.

Proper Use

All POP/IMAP passwords will have to comply with the new password policy. Please visit the Password Web site: <http://network.lanl.gov/passwords/>.

This site explains the new unclassified password policy and has a password checking utility to verify that the passwords you choose meet the new criteria. Each user was required to certify compliance on the Web site by September 24, 1999.

E-mail on the Internet is not secure. The common dictum is "Do not send any e-mail that you would not put on a postcard." Unencrypted e-mail should NOT be used for sensitive or private information (performance appraisals, etc.). There is no such thing as privacy on the Internet, including local mail. The new yellow (nee blue) protected/open network has access controlled by

a firewall. It was not designed to protect the confidentiality of data. Treat information in the yellow with the same regard you would on the green (unprotected/open) network.

Most e-mail clients have a method for configuring your mail to delay sending messages immediately. People who deal with sensitive and/or classified information, may contact their system administrator to set up this option. This option gives users a chance to review the material one more time before posting. If you are a user inclined to reacting without thinking before sending e-mail, you should consider this as well.

If you are forwarding or re-posting a message you have received be sure to give proper attribution and do not change the wording. Ask permission of the sender first if the message was personal or you are re-posting to a group.

Never send chain letters via e-mail. They are forbidden on the Internet.

A good rule of thumb to follow is to be conservative in what you send and liberal in what you receive. Do NOT respond to flames or spam. Remember that while you use an address identified by the "lanl.gov" domain, you represent the Laboratory in your conduct and words.

E-mail messages should always have an appropriate subject heading.

Make things easy for the recipient. Many mail clients strip out header information. Use a signature file at the end of your message with contact information in it. Avoid using HTML-formatted signatures. These signatures have embedded tags that wreck havoc with mailers.



Know whom you are addressing. If you are sending to large e-mail distribution lists like "cic-all," keep the message short and appropriate. (Plain text is best!)

If you receive anything that is questionable or illegal, forward it to the S Division Special Project Office.

Use mixed case in e-mail. Using all caps is considered to be shouting.

Send text messages only, unless they are Multipurpose Internet Mail Extensions (MIME) attachments. Many people, for instance, cannot read encoded Microsoft documents.

Be aware of the message size. E-mail is not appropriate for movies, software, and large PowerPoint documents. Those cute Santas dancing with hula hoops can reach megabyte proportions and when sent to large lists can severely impact e-mail services and cause "denial of service" to users on the server.

Don't send large amounts of unsolicited information.

Limit line length to fewer than 65 characters and end a line with a carriage return. Never assume that all mail readers will be able to interpret line wrapping.

Use mailing lists where appropriate. Large header files with hundreds of addresses are hard to read and do not make use of list software for grouped distribution to forwarding servers.

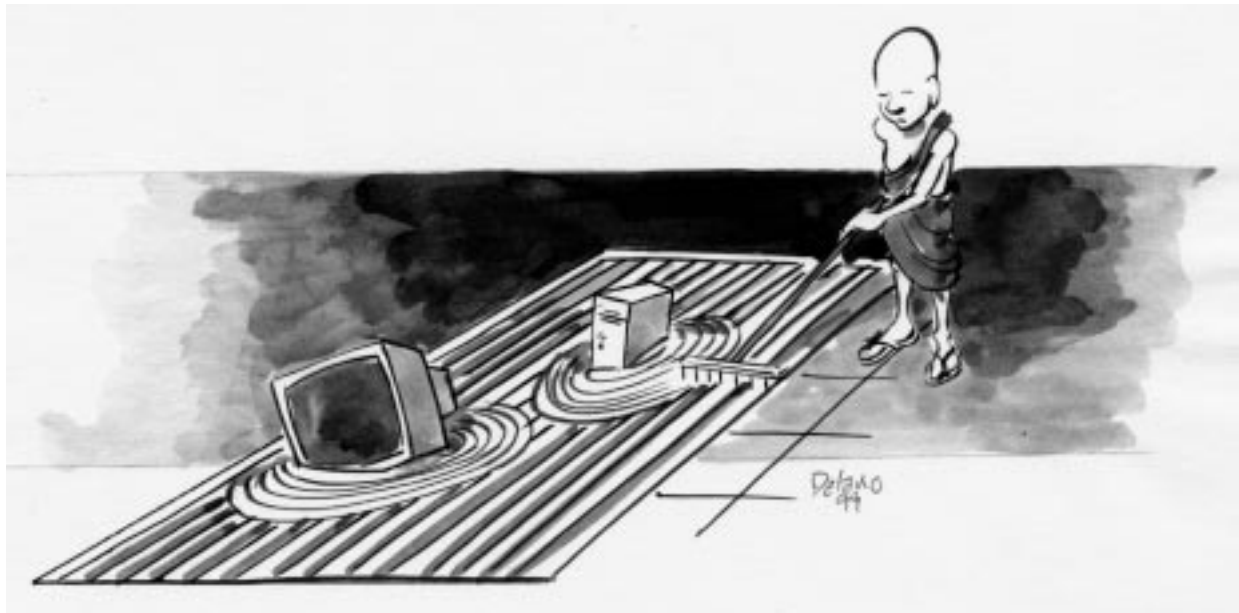
Spam (or unsolicited e-mail) is the biggest abuse of mail lists. All senders should take care that they are not the source of junk e-mail.

E-mail is subject to forgery and spoofing (assuming other identities or using pseudonyms). Be careful that you do not assume too much when you receive a preposterous message. For example, if you receive e-mail from your group leader awarding you a 50% raise, you might want to verify the e-mail before you buy a new car.

Records Retention

CIC does not encourage storage of e-mail on the mail servers. Servers can be compromised and e-mail clients do not perform efficiently when large folders are used. The place to keep stored e-mail is on your desktop with appropriate backup there. The CIC-5 e-mail servers are backed up nightly for system and service files and most e-mail will not be included in these backups. Operation logs are now archived for an indefinite period. If you lose a disk on your desktop, chances are good that your e-mail is lost unless you have backed it up locally.

For questions, comments, or further suggestions, please send e-mail to <mail-manager@lanl.gov> or call the CIC-6 Helpdesk at (505) 665-4444 ext. 851.



Data Analysis in the Advanced Surveillance Technology Initiative: AMISS Testbed

by Diana Mayer Orrick, Staff Research Assistant, CIC-15 Advanced Database & Information Technology

Data analysis and data mining involve the extraction of useful knowledge from large datasets and are considered steps in the process of knowledge discovery in databases (KDD). Automated data collection and tools have produced data repositories that contain gigabyte- and even terabyte-sized databases. Previously, human experts familiar with the application of the particular type of data would perform in-depth analysis, but the size of current data repositories expands the time that this would require beyond the capabilities of human analysts.

The approach to knowledge discovery within any data repository begins with consideration of the different data sources and the integration of the data into an easily accessible format.

Consideration must be given to the data analysis that is to be performed. Choices of appropriate software tools must involve weighing all analytical needs.

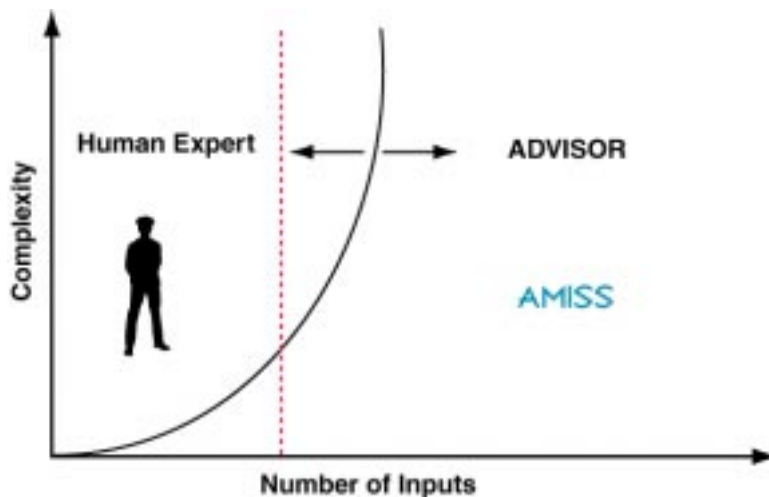


Fig. 1. AMISS expands beyond human capabilities: AMISS capabilities allow knowledge discovery from massive quantities of complex data from various sources (e.g., video, biometrics, radiation, etc).

Introduction

The Adaptive Multisensor Integrated Security System (AMISS) is a testbed that has been developed at LANL as a part of the Advanced Surveillance Technology initiative to identify potential threats and anomalous situations within nuclear facilities. The initiative is owned by Safeguards Systems (NIS-7), with task support

from Florida State University, the University of New Mexico, CIC-15, and Applied Theoretical and Computational Physics (X) Division, and Theoretical (T) Division. The AMISS testbed uses several different artificial intelligence (AI) technologies to provide a level of reasoning that can ascertain appropriate actions that increase facility security, safeguards, and safety. The system differs from ordinary security



Fig. 2. AMISS core competencies.

systems in that it is designed to operate in areas that have significant ongoing activity. It specifically targets actions that could be considered threatening to the facility and uses pattern recognition to distinguish these potentially threatening actions and behavior from normal behavior. Its primary functions are shape recognition, anomalous path tracking, pattern recognition, nuclear material detection and tracking, and object tracking. Various types of sensors are used in the system, in which some data are captured and analyzed in real time in addition to being archived, and the testbed allows for rapid inclusion of alternative sensors to address alternative surveillance challenges. Data are also stored in a database where they can subsequently be analyzed using data-mining techniques.

The data, both text-based and those held in relational databases, require multi-data-format consideration. The goal of the data-mining process used in AMISS is to detect deviation from sequential and time-series datasets. Other methods may be employed during experimentation.

Currently, data preprocessing is underway, leading toward final data analysis. The process of KDD begins with the database source. Raw data are integrated and cleansed of missing or incorrect data in preparation for data warehousing. From the warehouse format, task-relevant data are mined for particular pattern evaluation. Analysis of the patterns can lead to new mining or derives new knowledge from the datasets examined.

Further results of the data-mining project developed for AMISS will be published at a future date when the analysis stage of the work has been completed.

Data Format and Storage

The approach to knowledge discovery within a data repository begins with consideration of the different data sources and the integration of the data into an easily accessible format. Data might be derived from sources that are relational, transactional, object-oriented, spatial, time-series, text, or multimedia. One way to accomplish this integration is to create or build a data warehouse.

A data warehouse provides multidimensional access through a single database format to historical data that have been summarized and consolidated from the records of one or more operational databases. Once construction of a data warehouse is complete, data-mining tools are used to extract information.

Careful consideration of data storage and integration greatly promotes the success of a data-mining endeavor. In addition, the data-warehouse process should provide scalability, so that as the warehouse size increases there is no degradation in performance or requirement for modification of the applications used. Scalability considerations should include the amount of data, the complexity of queries of users, and the number of concurrent users who will access the warehouse.

Additional challenges to successful knowledge discovery include high dimensionality of databases and the danger of overfitting for a particular model. Data sources that change or are nonstationary and data sources with missing and/or noisy data present further problems. The need for better methods to understand the complexity of relationships between fields and for better visual representations of patterns for improved understanding will be included in future development of successful data-mining methods and tools.

Data-Mining Techniques

Consideration must be given to the function of the data mining to be performed. Association mining discovers rules for hierarchies within a dataset by identifying implications or inferences between items in a dataset. Classification on attributes within a dataset creates new classes and derives better descriptions of the information the new classes contain.

Clustering of records into subsets of homogeneous groups of records assists in identifying areas with high density within a large dataset. This method assists in solving problems of scalability and interpretation of highly dimensional data.

Some of the techniques applied to classification and clustering include neural networks, genetic algorithms, and statistical algorithms. At the present time, neural networks and statistical approaches—such as Maximum-entropy-driven Bayesian methods—are being investigated as a possible means to analyze AMISS sensor data.

Identifying subsequence patterns within sequential data leads to opportunities for prediction-based studies. The problem is to locate within the dataset all sequential patterns with a user-specified minimum support, where support of a sequential pattern is the percentage of data sequences that contain the pattern. Often, the sequential data are examined in the interest of trend analysis as a time series for patterns of cycles or periods. Current work in time-series analysis seeks to identify partial periodic patterns as well as the full-cycle periodicity searches of previous methods.

Another area of data-mining methods identifies deviation in patterns found. These anomalies may lead to identification of problem areas or new opportunities. For data mining, one or many of these methods might be necessary. Through the application of appropriate algorithms and techniques, the data-mining step in the knowledge-discovery process can assist the user to identify interesting aspects of items within a database.

Software

Considerations for choosing appropriate software tools include weighing all previous considerations of data format(s), data storage and access, and the data-mining method(s) to be employed. The software should provide flexibility, considering current and future data sources, and should support multiple data formats. The processes of data storage and data access should be compatible with the software choice. Scalability and complexity issues regarding performance also should be reviewed. Multifeatured software that supports multiple data-mining methods provides the broadest potential for yielding meaningful data-mining results. In addition, a software tool that produces visualization of data-mining results with user-directed interactive options can provide greater acceptance and ease of use.

Because of its affordability, visualization tools, and multifeature and multi-data-format support, the software initially chosen for this data-mining project was DBMiner, developed at Simon Fraser University in British Columbia, Canada. The software proved to be troublesome to implement and failed to provide interesting information for the AMISS data examined. DBMiner intends to release a professional edition of the software later this year, and at that time many of the shortcomings of the educational release may have been resolved.

The Preprocessing Effort

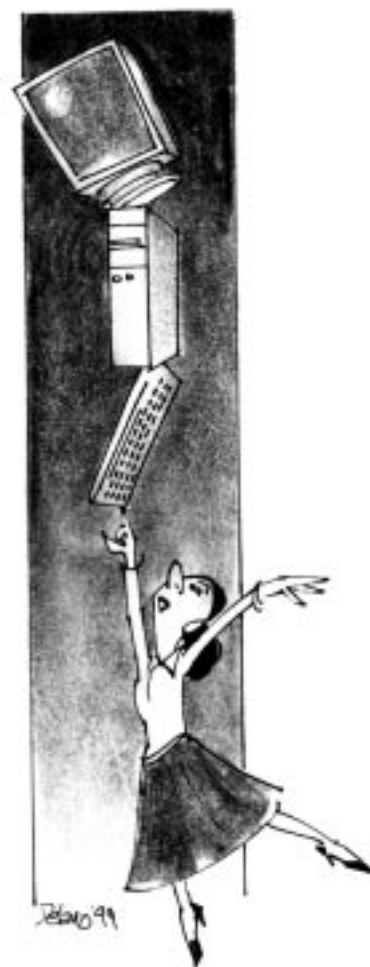
During summer employment at LANL, the author focused on preprocessing data for analysis. The aim was to provide easily accessible formats of the raw data sources for examination through user-developed computational methods of data analysis. The data-preprocessing or data-preparation stage in the knowledge-discovery process may often require up to 70% of the total effort expended in successful data analysis.

Because much of the raw data exist in text-based log files, a text-manipulation tool was needed to provide fast, efficient processing of large data-log files. ActivePerl™, developed by the Active State Tool Corporation, provides a Microsoft® WinNT application for using perl and is available free of charge. More information on ActivePerl is available at <http://www.activestate.com/ActivePerl/default.htm>. Additional information and downloadable modules for perl can be found at <http://www.perl.com/CPAN/>. The perl programming language is known for its text-manipulation capabilities, and the WinNT version of the compiler was chosen to retain the OS used primarily for data storage in AMISS.

ActivePerl has provided an easy environment for the fast development of programs designed to format the raw data sources into data-specific format files. Some of the programs developed to date provide multisensor to single-sensor-type data, single-sensor-type to individual-sensor data, time sequencing of multiple-sensor data, and extraction of firing times between the different sets of data formats. Additional program development focuses on other data fields provided in the sensor-feedback log files.

Conclusion

In the future, data mining and data analysis will identify, from the preprocessed information, patterns that can be identified as "normal." With a definition of normal established, anomalous patterns will be sought to determine nonnormal activity from archived, long-term data from AMISS. Results of the data-analysis effort will be included in the MS thesis of the author. More information on the AMISS project can be found at <http://nis-www.lanl.gov/nis-projects/amiss/>.



Transaction Integrity and Validity in MC&A Databases

by the Database Assessment

Project Team: Jay M. Armstrong (S-4), Paula L. Knepper (NIS-7), Cindy J. Mills (NMT-2), Kimberlyn Mousseau (CIC-15), Sharon L. Seitz (NIS-7), Rena Whiteson (NIS-7), and Sandra E. Wilson (NMT-4)

The Database Assessment Project is being developed collaboratively between NIS-7 and TA-55 Groups (specifically NMT-4 and S-4) to improve the accuracy of the information stored and reported about the nuclear materials at LANL. Because Department of Energy (DOE) sites use computerized materials control and accountability (MC&A) systems to track and account for nuclear materials, it is important to eliminate as many errors as possible in the database used by the system. To determine the accuracy of the information, it is first important to consider the activities required to accurately track the materials and the problems associated with capturing the information about them. Though process-accountability procedures are defined and implemented at LANL, they cannot address every possible situation that may occur when materials are taken from one form to another, repackaged, and/or transferred to different locations. Once tracking information has been collected, processes for assessing data quality, how those processes might be improved, and what can be done to automate them must be considered. The expert system work attempts to detect not only errors within the transaction but also anomalies within the activities that created the transaction.

The objective of the expert system is to automatically determine whether MC&A transactions are correct.

Because the diverse nature and large volume of data make manual verification difficult and tedious, automated error-detection capabilities are highly desirable. Past work in this area has focused on historical data, periodically assessing it for specific errors. The focus of the current effort is to guide, suggest, warn, or prevent material handlers from doing anything unusual, violating procedures, or deviating from specific guidelines by analyzing transactions in both real time and batch modes. This will be accomplished by comparing transactions against process accountability flow diagrams (PAFDs). These diagrams graphically depict correct procedures for processing nuclear material. The objective will also be satisfied by "capturing knowledge" that is currently available only by talking with domain experts. Combining expert system and advanced database technology will not only improve the accuracy of the data contained within the MC&A system but will also reduce the manpower currently required to assure the accuracy of the information. Figure 1 gives a graphic representation of the current MC&A process flow.

MC&A Systems

Materials control and accountability systems are developed as part of the safeguards program as required by the DOE Order 5633. Their main objective is to store information about the location and status of nuclear materials. Additional information tracked by these systems includes measurements and measurement instruments, measurement timestamps, and materials-related processes that are performed within a facility. These systems must ensure the accuracy of the information about the quantities of nuclear materials being tracked, including those that have been shipped, received, or otherwise removed from an area. Material-type codes and project numbers as published by DOE are required for all transactions and must also be stored by the system.

The Database Assessment Project deals with transactions produced throughout the Laboratory but concentrates mostly on transactions involving operations at the vault at the Los Alamos Plutonium Facility at Technical Area (TA) 55. This information is important for programmatic

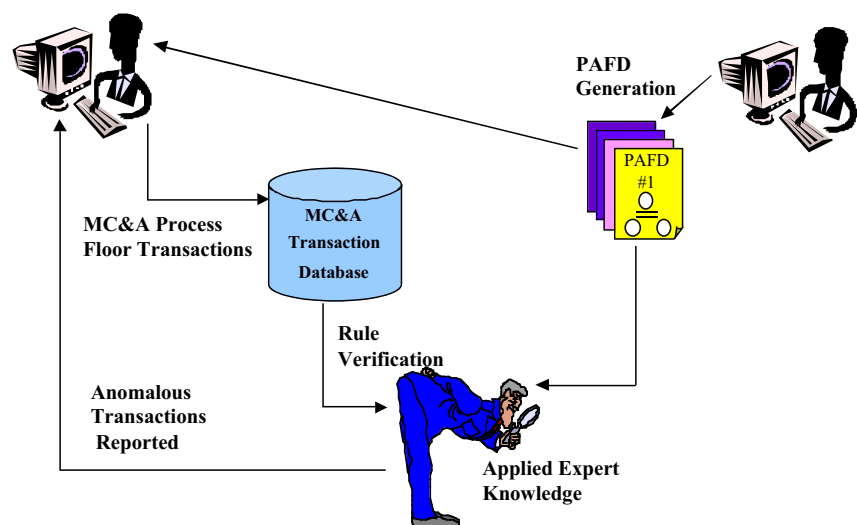


Fig. 1. A graphic representation of the current MC&A process flow.

reporting required by DOE Headquarters, Operational Offices, and, in some cases, Area Offices. To fully satisfy these reporting requirements, it is important to ensure that the information is correct. However, exceptions may occur while tracking nuclear materials that make anomaly and error detection of the information difficult. Instrument calibrations may vary and variances may exist that override acceptable procedures. Discrepancies may also exist among what actually occurs while taking a measurement, what the technician reports, and what was supposed to have occurred. Frequently, measurement information is entered by hand resulting in errors occurring during data entry. And, in many situations, only domain experts know whether a transaction contains an error.

Past work has been successful in automatically detecting some errors and anomalies in existing nuclear-material databases. Other anomalies and errors are detected manually. However, manual detection requires a large investment of time by MC&A experts. The intention of this work is to extend on what has already been accomplished by capturing additional expert knowledge in an expert system. This system will be developed to evaluate transactions against the rules defined by the expert to determine if a transaction contains anomalous, erroneous, or unusual information. In the future, the expert system will be enhanced to compare transactions against the process information defined in PAFDs.

Process Accountability Flow Diagrams

PAFDs are graphical representations of the correct processing of nuclear material. They are required for any special nuclear material (SNM) operation in which there is the potential for gain or loss of material. PAFDs are



developed and maintained at LANL to provide guidance for processing nuclear materials and to describe when MC&A measurements must be taken. PAFDs contain the feed items of a subprocess, i.e., the rules that apply to feed measurement codes and the outputs of products produced by each feed. When an operator has been notified that an item has arrived at the vault, before placing the material into the vault, the operator locates the applicable PAFD and verifies that all correct procedures have been followed and completed. Some, not all, PAFDs are used by the operators to specify lot identification (LOTID) changes that allow the operator to know where the material is in terms of the overall process. That is, as material flows through the process, different LOTIDs are assigned to the material. The PAFD is used to identify the points in the process at which new LOTIDs are assigned. The operator can get a listing from Material Accountability & Safeguards System (MASS) of the material in his/her process and know exactly how much material is currently at any point by referring to the LOTID.

In order to automate the analysis of MC&A transactions against the PAFD, it was necessary to convert existing PAFDs from PowerPoint files into an electronic format that is easy to maintain and accessible to the expert system.

After evaluating several commercial relational-database and graphing-software packages, VISIO Enterprise™ was selected. VISIO Enterprise is a software package that enables design and management of the drawings. VISIO Enterprise allows a drawing to be stored not only as a drawing but also as a database. The drawing defines the information stored in the database and the database can be stored in any open-database-connectivity- (ODBC-) compliant format.

Using the VISIO Enterprise package, various tools were created to assist in generating PAFDs. These tools were created to provide a user-friendly method for creating the PAFDs, make them easy to modify and maintain, and allow access to the information contained within the PAFD by various other programs. A stencil was created that

provided a user-defined toolbox and items were selected to represent the elements necessary to generate a PAFD. Once the PAFD is created in VISIO, the data contained within the drawing can be exported to a database using a built-in macro where the information to be stored in the database is defined by the user during the export process. Additionally, VISIO provides the ability to modify either the database or the drawing and to pass the changed data between the two. This feature improves the maintainability of the PAFDs.

Transaction Analysis Using the Anomaly Detection Expert System

For the expert system to be effective, it was necessary to select an expert system tool that could satisfy several high-level requirements. The tool had to be robust enough to capture the knowledge and reasoning known by domain experts to flag suspicious MC&A transactions. Because of budget constraints, it was important that it be cost-effective and easy to learn. It also had to interface with an ACCESS database to retrieve PAFD information so that transactions could be compared against process flows. Lastly, it was important to ensure that the expert system could easily transition; when data exchange models are updated and legacy MC&A systems are transitioned to modern storage systems, the expert system must be capable of transitioning with them. Work in this area began by evaluating off-the-shelf tools that could satisfy each of these requirements.

Acquire® was selected as the expert system tool because it is intended for rapid and cost-effective development of an expert-system knowledge base. It was designed for the novice user and requires no special training in expert-system technology. It is based upon two fundamental concepts: assisting the expert rather than the knowledge

engineer, and recognition of specific patterns rather than the application of general rules to specific situations. Acquire provides the ability to effectively organize an expert's knowledge by structuring the information into meaningful patterns that are easy to observe. It also provides a software development kit that allows the knowledge base to be embedded into a Visual Basic or Visual C++ application, hereafter referred to as the expert system container. Through the container, it is possible to connect to any ACCESS database.

The expert-system container is being developed using Microsoft's Visual C++ and has three primary functions. First, it is responsible for providing the user interfaces that are necessary to obtain the information that will be passed to the expert system. Two interfaces exist: the first receives transaction information as defined in the current MC&A system and provided the ability to analyze transactions in real time. The second extracts existing transactions from an ACCESS database and analyzes them in batch mode. In the former case, if a transaction is found to violate a rule defined in the expert system, the end user is notified and then presented with an option to continue or correct the error and resubmit the transaction for analysis. In the latter case, any transaction containing an error is written to an error database for human intervention.

The second primary function of the expert-system container is to access the exported information about a PAFD drawing from the ACCESS database produced by VISIO. This information will be used to compare the values in the transaction to qualify the information against predefined processes. Finally, the container application is responsible for passing transaction information to the knowledge base and communicating any messages sent back by the expert system to the end user.

To obtain the rules for the expert system, an MC&A domain expert was consulted. Information was collected that would determine the relationship among a few of the attributes belonging to a transaction. Once relationships were defined, a list of every possible combination that the attribute values could assume was constructed and defined in Acquire. For the initial set of rules, combinations of LOTID and IDES were compared. IDES codes describe all nuclear material items residing in the Laboratory inventory. Each IDES code consists of a four-character alphanumeric field. The first character identifies the major category under which an item belongs. The second and third characters provide further detail of the major category by describing the material's modifier, e.g., dioxide, carbide, tetrafluoride, etc. The last character is used to identify some specific characteristics of an item, e.g., a major component or alloy. It can also be used to provide some other pertinent piece of information about the item, such as high purity.

LOTIDs are assigned at different steps during material processing and contain substrings that correspond to the meaning of an IDES code. For instance, if *MET* is a substring of LOTID, the first character of IDES must be equal to *M* indicating that the material falls under the major category of Metal. Output values are assigned to each combination of IDES and LOTID where the value set consists of *Accepted*, *Rejected*, or *UNKNOWN*. *Accepted* is assigned to valid combinations, *Rejected* to invalid combinations, and *UNKNOWN* is assigned if one of the attributes contains a null value. Output values are then passed back to the container application and communicated to the end user.

Even though a transaction is rejected by the expert system, it may not actually contain anomalous or erroneous information. Situations may arise where, since development of the expert system, values or rules have changed, per the Accounting Office. Other transactions may appear anomalous because information about newly processed items has not been correctly updated in the MC&A system. In some cases it is required that the Accounting Office review the transaction before authorization is granted to update the transaction. It is important that the expert system notify the material handler when an apparently anomalous transaction exists, but it must also allow the user to make the final determination.

Future Aspirations

In the future, the expert system will be modified and enhanced to satisfy several additional requirements. Transactions will be compared against the PAFDs. The expert system will be incorporated into the Integrated Nuclear Materials Information System (INMIS), a system currently under development to replace the existing MC&A system at TA-55. Once the expert system is embedded into the INMIS system, transactions will be analyzed in real time. It is also the intention to allow batch-mode analysis via an Intranet. Further, capabilities developed for this system will be used in the Adaptive Multisensor Integrated Security System (AMISS) project. The objectives of AMISS are to provide continuous, adaptive, real-time detection and categorization of all activity by tracking personnel and material at LANL. AMISS is also being developed to demonstrate state-of-science adaptive technology to determine potential threats and anomalous situations. The expert system was designed using object-oriented methods and will easily transition to benefit other projects throughout LANL.

In addition, the system will be enhanced to automate knowledge acquisition. This would allow experts to add or modify the information in the expert system without help from knowledge engineers. Because information corresponding to an MC&A transaction can change, domain experts must have the ability to reflect those changes in the expert system. Allowing the expert to make the changes rather than relying on the knowledge engineer will reduce the amount of time it would otherwise take to reflect MC&A transaction changes in the system.

Bottom Line Impact

Bottom line, the Database Assessment Project is the result of longstanding, ongoing collaborations between NIS-7 and TA-55 Groups and is being developed to automate the detection of erroneous and anomalous MC&A transactions. Current work focuses on capturing the information known by domain experts whereas in the future, the expert system will compare transactions against the PAFDs and will analyze them in real time. The expert

system will be embedded in the INMIS system and will allow experts to change the knowledge within the expert system without help from knowledge engineers. Once the expert system has been developed and is successful at TA-55, other areas across the Laboratory where this system might be modified and enhanced to provide further benefit will be defined.



Frameworks Are Models, Too!

by John H. Hall, Technical Staff Member, X-10: Code Group B

The Tecolote framework is a model of a typical application. In fact, there is nothing in Tecolote's design that is specific to any particular type of application. You could use it as the basis for a word processor as easily as we are using it for physics simulations. Our customer's area of concentration is running batch applications that cycle over time steps in the problem domain and that is the topic of the rest of our discussion.

Tecolote 2.3 (our current production framework) uses a sequence of setup actions followed by an event loop. In physics simulations, for example, we normally loop over simulation time. In the upcoming version Tecolote 3.0, we can actually describe multiple threads of execution, branching, and other computer science constructs because we are using a general-purpose functional language as our scripting language. Figure 1 below shows the sequence of events in a typical application built using Tecolote 2.3.

The power of this simplified Tecolote 2.3 model of an application might not be immediately evident. In fact, almost every modern application can be described using this abstraction. The

only difficulty arises when you attempt to implement a multithreaded code with multiple independent threads that execute different algorithms simultaneously on different pieces of data (MIMD). However, Tecolote 3.0's additional flexibility will allow even this abstraction to be fully described.

Because the Tecolote 2.3 model is able to describe most applications, it is no longer necessary to use separate applications for every conceivable purpose. Instead, you can build a single collection of components and—by simply using a different input file—you choose the set of objects to be built in memory; these components, when run, comprise the actual end-user application.

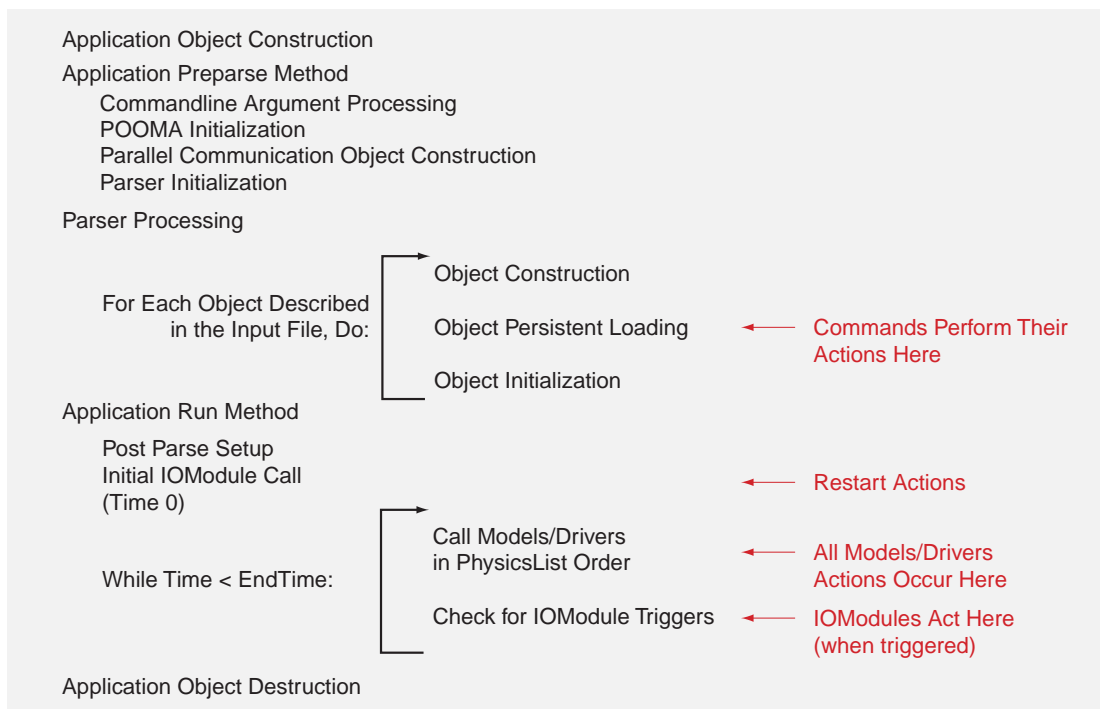


Fig. 1. Sequential description of actions for an application created using Tecolote 2.3.

Navigating in Tecolote

Tecolote Organization

The vast majority of programmers can use a simple flow chart to determine where a new model or command should be added to the framework. Follow the chart below (Fig. 2) until a leaf box is reached. That box will contain the name of the object you want to create.

Tecolote Object Descriptions

The **objects** shown above are examples of a special family of **objects** that model the concept known as "Tecolote Traits." In practice, a small number of traits allow almost all classes to be described by using a simple abstraction of "construction, persistent loading, initialization, and evaluation."

We have developed the ability to handle four categories of objects:

- basic/intrinsic types such as int or float,
- non-Tecolote-aware classes such as those found in third-party libraries,
- Tecolote-aware classes, and
- Tecolote-aware wrapper classes that extend the functionality of non-Tecolote-aware classes. (We use the external polymorphism technique to wrap some important non-Tecolote-aware classes, allowing them to be full participants in the Tecolote Framework; for example, the STL vector class and the POOMA Field class have both been extended in this way).

When a new model is introduced into the framework, it is best to make it a full participant by satisfying the few requirements imposed by the framework. Stub member functions that satisfy these requirements are automatically provided through inheritance by the base classes shown in Figure 2, allowing you the freedom to concentrate on new functionality. However, if you choose to create a new hierarchy, you must implement the requirements yourself. Figure 3 demonstrates how required functionality is established for a Tecolote Component in the form of a class description.

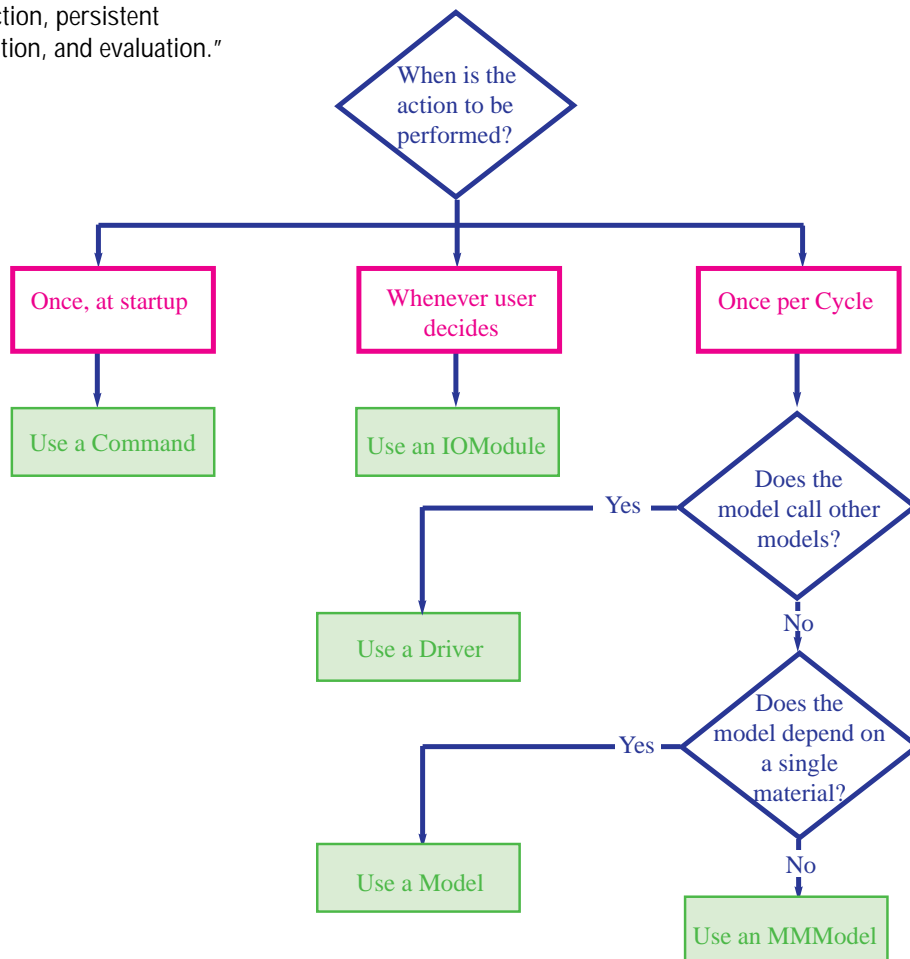


Fig. 2. Decision tree for choosing a Tecolote object.

```

template<class T>
class TecObject {
// A macro is required to tell Tecolote that this object has persistent data
    PERSISTENT_MEMBERS(TecObject)
// A constructor is required, which takes as its arguments a pointer to a DataDirectory
// and a const reference to a string
    TecObject(DataDirectory* dd, const string& name);
// The six C++ Default Methods should be listed; those being supplied
// by the compiler are commented out
//     Commenting out the default constructor implies a
//     compiler-supplied default constructor
//     TecObject();
//     We are supplying the destructor
    virtual ~ TecObject();
//     We are supplying the copy constructor (which is required in Tecolote 2.3x,
//     but not in Tecolote 3.x)
    TecObject(const TecObject<T>& c);
//     We are supplying the assignment operator
    TecObject<T>& operator=(const TecObject<T>& c);
//     The compiler will provide the address of operators
//     TecObject<T>* operator&();
//     const TecObject<T>* operator&() const;
// Methods
// An initialize method is required
    void initialize(void);
// A getDataDir method is required, although it is usually inherited
    virtual DataDirectory* getDataDir(void);

// Operators
private:
//     Type                Member_Data;        // Description
//     DataDirectory        *dd;                // Entry point to DataDirectory Hierarchy
};

```

Fig. 3. Example of a Tecolote-aware class declaration.

The five Tecolote objects depicted in the decision tree in Figure 2 are further defined below.

Tecolote Commands

A command object performs one function and then disappears. When called from the input file, the command class goes through the normal process of construction, persistent loading, and initialization. The command object performs its function in the method called “initialize;” for example, setting up the volume fractions and velocities for an analytical problem, creating a special field for a test problem, or rotating a 2-D object to give a 3-D object. After performing this function, the command object is deleted. Only side-effects manifested during the command object initialization persist.

Tecolote IOModules

An IOModule performs its desired actions only when requested by the user. During each cycle, the Tecolote Framework compares the current problem time to the next trigger time specified for each IOModule. Only the IOModules that need to take action then are sent a message to perform their action. Because the time abstraction that determines whether to call an IOModule belongs to the Tecolote framework, this mechanism only had to be written once to be available for every IOModule.

Tecolote Drivers

A driver simply calls other models (or drivers). A simple example is an EOS driver, which sequentially finds every material in the problem, asks if it has an EOS, sets up the environment for calling the EOS, and then makes the actual call. Sometimes the driver will need to contain some calculations that tie the individual models together. If

the simulation calls for a predictor-corrector model, a driver that understands the predictor-corrector sequence will also be required.

Tecolote Models

A Model is dependent on a *single* material. Models can be either time-dependent or time-independent. They can also be arbitrarily complex; however, it is a good design decision to factorize a Model (or any Tecolote component really) to perform a single task. Several of the most commonly used Models in the Framework are Equation of State Models (EOS) and the Ramp Model treatment.

Tecolote MMModels

The MM in MMModel refers to multiple materials. There are two categories of MMModel. If the model explicitly needs to iterate over each material and extract information to perform its duties, it is clearly an MMModel. However, another possibility is that the model simply doesn't rely on any single material. Examples of MMModels include artificial viscosity and timestep controllers; as a special case, Drivers are derived from MMModels.

Summary

We have shown the execution model underlying the Tecolote 2.3 Framework. New programmers could now navigate to the appropriate place to add their code. Of course, this brief introduction doesn't begin to touch on all the helper services available within the Framework, and to do any effective physics simulation would require familiarity with the POOMA framework. The Tecolote Team has been developing techniques that allow code to be written once and used for 1-, 2-, or 3-D simulations without any changes to the source code. These design patterns, helper services, and other enhancements will be described in depth in future issues of BITS.

Commands and IOModules share a troublesome aspect. Because they can be asked to perform their tasks at any time, they cannot assume that the entire application infrastructure has been constructed and is ready for use. Consequently, these components must be able to ensure that any conditions required for their operation are met—even if the operation requires object construction—before they try to perform their tasks.

Lecture Review: Automated OS Install HOWTO: Linux Cluster via Gigabit Ethernet

*by Susan Coghlan and Ron
Minnich, Technical Staff Members,
Advanced Computing Laboratory*

Large clusters of computers are becoming a common way to deliver high performance computing for a fraction of the cost of more traditional supercomputers. However, installing and upgrading the OS on a large cluster without automated tools can be a daunting task. For example, it can take up to a week to install the OS on a 64-node cluster using traditional interactive user interfaces.

Using Red Hat Linux 6.0 Kickstart Process

On September 2, 1999, Aaron Marks from the University of Pennsylvania presented a talk illustrating how to use the Red Hat®, Inc. Linux 6.0 kickstart process to install Linux in an automated, configurable, distributed method quickly across a large cluster of nodes. This process can be particularly difficult when the nodes do not have monitors ("headless" nodes) and the drivers for the network interface cards (NICs) are not included on the standard kickstart disks. In Marks' case, he had a 20-node cluster with no keyboards or monitors and nonstandard Gb/sec NICs.

Marks built custom kickstart CD-ROMs that were used to install the distribution from an NFS mounted image, but it is also possible to use a kickstart floppy (rather than a CD-ROM) and to install the distribution from an image on a local CD-ROM. Marks pointed out that he prefers CD-ROM to floppy drives because they are faster and more reliable.

Using kickstart, a full install of Red Hat Linux 6.0 can easily be done in parallel across all nodes in a cluster with minimal manual intervention necessary. Marks' method does require physical intervention at one stage in the process where the kickstart disk must be ejected during one of the boot sequences.

There are three stages to the installation. The bulk of the work is done in Stage 1, with the initial boot using the kickstart kernel, the actual install of Red Hat Linux 6.0, and the copy of the latest kernel into the /boot directory of the new 6.0 install. The kickstart process initiates a reboot at the end of Stage 1. After the system has fully booted, the new kernel copied into /boot during Stage 1 is loaded using the standard Linux boot loader (lilo), and then the machine is rebooted one last time, completing Stage 2. The third and final stage of the kickstart process is to reconfigure the network configuration files using the information obtained from the dhcp/bootp server. Complete details for each stage can be found in the HOWTO at <http://www.cis.upenn.edu/~ajmarks/>.

This article will provide a short summary of the various stages and the issues within each stage.

Stage 1

For Stage 1, you will need to build a custom kickstart disk with a kernel that includes the necessary drivers for your particular hardware, and a modified SYSINUX configuration file (syslinux.cfg) that passes special flags to the kernel (in particular, serial support is enabled that allows noninteractive progress monitoring of the headless nodes). To do this, you modify the bootnet.img from the RHL distribution (located in the Images directory).

If you are building a kickstart floppy, be careful to keep the kernel as small as possible so that it will fit on the disk; however, if you are building a CD-ROM it is unlikely that you will run into any such space limitations. It is necessary to include serial support and also to modify the include/linux/version.h file to define UTA_RELEASE as a specific string (2.2.2-14BOOT for the 6.0 release). The UTA_RELEASE string must be exact or the kickstart process will die. You could also try to match version information everywhere there is a loadable module, but this is much more complicated and precludes using the vanilla Red Hat distribution tree. Detailed instructions for building the kernel and the kickstart disk are provided in the HOWTO referenced above.

You will need the following network services: DHCP/BOOTP (network configuration information), DNS (hostname resolution), and NFS (kickstart configuration file and OS distribution). It is perhaps easiest to set up a dedicated kickstart server to provide the DHCP/BOOTP and the NFS services.

A kickstart configuration file is required for each node; however, symbolic links may be used if you wish to have multiple nodes use the same file. The configuration file is complex and an example file is given in the HOWTO.

Marks pointed out some hidden kickstart directives: one that can be used to cause kickstart to immediately reboot after the installation process and one to turn off probes for a mouse. Without the reboot directive, kickstart will wait for a RETURN key press, making an unattended, headless, keyboardless install impossible.

Using NFS to do the actual install of the distribution instead of loading from a local CD-ROM allows you to customize the distribution and to keep it current as updates to packages become available. According to timing runs that Marks performed, the difference between loading over the network and loading from local CD-ROM was negligible, no doubt a result of using the GigE NICs. Marks reported that on an earlier cluster with 100baseT interfaces using the local CD-ROM was much faster.

Modifications are made to the `rc.local` startup script during the `postinstall` section of the `kickstart` script. These modifications initiate Stage 2 upon reboot of the node at the end of Stage 1.

Stage 2

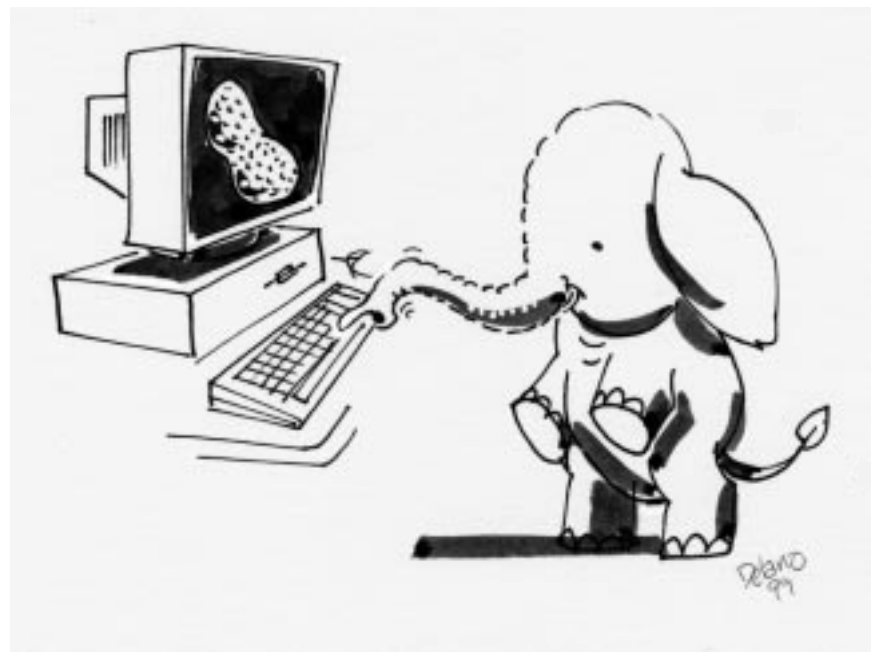
Stage 2 is a simple stage: the new kernel is linked into the standard name, `lilo` is run, and the node is rebooted. This is where physical intervention is required, because the `kickstart` CD-ROMs and/or floppies must be removed before the node starts final boot. If the media is in the node during the boot sequence, the node would start over with Stage 1.

After this final reboot, the machine will come up with the latest kernel, support for the gigE NIC, the unique network configuration for the node, and serial access enabled through COM1.

Using his customized Red Hat `kickstart` disks, Marks was able to do a full Red Hat Linux 6.0 install on 20 nodes in less than 25 minutes. His modifications allow the completely automated, fully configurable install or reinitialization of the Linux distribution in 9.5 minutes per node just by booting with the `kickstart` CD-ROM loaded. (The cluster was split into groups of nodes and installs done serially across the groups because of the need to remove CD-ROMS during the final reboot in Stage 2.)

For Further Information

Marks is in the Ph.D. program at the University of Pennsylvania (Philadelphia) and is also on the Research Staff at the Sarnoff Corporation. He helped build several clusters while at Sarnoff, including the 161-node, 288-processor Cyclone cluster. Cyclone is a mixed Alpha/Pentium cluster. Marks' earlier work at the Supercomputing Research Center in Bowie, MD, includes a compiler for the Cray 3/SSS system, as well as parallel C compilers for clusters, FPGA-based machines, and Processor-in-Memory (PIM) systems. Marks can be reached at ajmarks@seas.upen.edu.



A New Look and Feel to xxx.lanl.gov

by Frances Knudson, Library Without Walls Team Member, CIC-14 Research Library

e-Print arXiv is a new interface to xxx.lanl.gov (or arXiv.org), the fully automated electronic archive and distribution server for research papers hosted at LANL. The archive started in August 1991 and covers primarily physics and related disciplines, as well as mathematics, nonlinear sciences, computational linguistics, and neuroscience. The new interface offers searching across all the archives of xxx.lanl.gov, multiple indexes, limiting, and sorting of search results.

The development of this interface stems from two goals of the Research Library and Library Without Walls Project. The first goal is to integrate preprints more into the traditional library literature. The second goal is to provide a common interface to as many scientific/technical literature resources as possible.

The first step has been to create a search interface into the xxx.lanl.gov e-print archive that mirrors the SciSearch® at LANL search interface. Library customers should find the navigation and the searching techniques very familiar. This interface is currently also available for INSPEC at LANL, and the consolidation of interfaces will continue through the fall as we move other citation databases into the SciSearch® at LANL model.





The LANL Research Library offers a variety of training opportunities for the Laboratory community. Available sessions focus on specialized library databases and other electronic resources. A complete list of course offerings can be found at <http://lib-www.lanl.gov/libinfo/training.htm>. All sessions are available to individuals or groups, at the library or your site. Arrange for a session by contacting the Library, phone 7-4175 or e-mail library@lanl.gov. Library tours are available on a drop-in basis every Wednesday at 1:00 p.m.

Attention Application Developers!!!

Need User Guides?

Need Online Help?

CIC-1's Software Documentation Team can help you improve your product by developing user manuals, online help, quick reference cards, or other documentation to help your users learn the application. Contact Sheila Molony, 5-1585, or sheila.molony@lanl.gov, for more information.



Computer Training

The Customer Service Group (CIC-6) offers technical computer training (enterprise information applications, communications, office administration, and Web authoring) and advanced computer training (programming languages, system administration, and advanced applications). To register for a course access our Web page at <http://www.lanl.gov/internal/training/training.html>. Or from the LANL home page select the links: Training, Computer. For further information about technical computer training call (505) 667-9559, and for advanced technical computer training call (505) 667-9399.

Technical and Advanced Technical Computer Training Courses		
Communications	Office Skills 2000	Web Authoring and Browsing
<ul style="list-style-type: none"> • Eudora 4.x • Meeting Maker 5.0.3 	<ul style="list-style-type: none"> • Office Skills 2000–LANL • Office Skills 2000–Professional Development 	<ul style="list-style-type: none"> • Dreamweaver 2.0–MAC or PC • FrontPage 2000 • HTML Basics • HTML Intermediate
Enterprise Information Applications (EIA)	Other EIA Courses	System Administration Training
<ul style="list-style-type: none"> • Date Warehouse–Basics • Date Warehouse–EDS Reports • Data Warehouse–Passport • EDS —Basics • EDS–GUI • EDS–Training Plans • Infomaker • Invoice Approval System • Key/Core System • Purchase Card System • Procurement Desktop • Recharge • Time & Effort GUI • Travel Foreign GUI • Travel Domestic GUI • Web JIT 	<ul style="list-style-type: none"> • Financial Management Information System (FMIS) • Property Accounting, Inventory, and Reporting System (PAIRS) • Signature Authority System (SAS) • Secretarial/Contract Service (SE) • Salary Review System (SRS) • Directory Information System (DIS) • Automated Chemical Information System 	<ul style="list-style-type: none"> • IRIX (SGI) System Administration (Beginning) • IRIX (SGI) System Administration (Advanced) • IRIX (SGI) Network Administration • SGI Performance Evaluation and System Tuning • Solaris 7 System Administration • Solaris 7 Network Administration • Solaris 7 Server Administration • Unix and Widows NT Integration • Windows NT Workstation and Server • Windows NT Optimization and Troubleshooting • Windows NT Security
Programming Training	Application Training	ASCI
<ul style="list-style-type: none"> • C Programming (Beginning) • C Programming (Advanced) • C++ for Experienced C Programmers • ANSI/ISO C++ Programming Clinic (Advanced C++) • Distributed Objects Using Corba • Java Programming • Java Program Workshop • Distributed Programming with Java • Programming for Beginners Using Java • Object Technology–A Management Overview • Object-Oriented Analysis and Design • Perl Programming • Advanced Perl Programming with CGI • C-Shell Programming • Python for Scientists and Engineers 	<ul style="list-style-type: none"> • Advanced WWW Development • FrameMaker Basic and Advanced • Foundations of IDL Programming • IDL 5.0 Graphic Object Workshop • Mastering MFC Development Using Visual C++ 6 • Netscape Servers for Intranet Development • Origin 2000 Applications Programming and Optimization • Sendmail–Managing Internet Mail • C++ and the Unified Modeling Language • Sybase Fast Track to Adaptive Server Enterprise 11.5 (ASE) • Sybase Performance and Tuning for System 11 • Sybase SQL Server Administration • Unix (Beginning) • Unix (Advanced) • Visual Basic 6.0 Fundamentals • Visual C++ Windows Programming • Windows NT Security • Windows 2000 Training 	<ul style="list-style-type: none"> • Getting Started on ASCI Blue Mountain Systems • Running MPI on Blue Mountain Systems • Introduction to Totalview • LSF (Load Sharing Facility) • Introduction to HPSS (High Performance Storage System)
* You do not need an ICN password to use e-mail.		

Reader Feedback

Feedback helps us to provide a document that responds to the changing needs of its readership. If you have comments or questions about this publication, please let us hear from you. We have reserved the back of this form for that purpose. We also accept articles for publication that are of interest to our readers. Contact the managing editor for more information. This form is also used for new subscriptions, deletions, or changes. Instructions are on the back. If you prefer to contact us by e-mail, send your comments and/or subscription request to denise@lanl.gov.

Do Not Staple
Fold on This Line First



NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL

FIRST-CLASS MAIL PERMIT NO. 88 LOS ALAMOS NM

POSTAGE WILL BE PAID BY THE ADDRESSEE

MAIL STOP B251
ATTN: DENISE SESSIONS, MANAGING EDITOR
CUSTOMER SERVICE GROUP (CIC-6)
LOS ALAMOS NATIONAL LABORATORY
PO BOX 1663
LOS ALAMOS, NM 87545-9916



Do Not Staple, Seal with Tape
Fold Here

cut along dashed line

Feedback

[illegible]

New Subscription, Deletions, and Changes

BITS is published by Los Alamos National Laboratory. If you would like to be added to or deleted from our mailing list, please check the appropriate line, complete the form below, and mail us the form, or e-mail bitsupdate@lanl.gov

_____ Add my name to the BITS mail list.

_____ Delete my name from the BITS mailing list.

_____ Change my name/address as indicated below.

Name	Date	
Address	Mail Stop	
Group	Organization	
City	State	Zip
Phone	Number of copies	

1999 12-Month Index

For a more complete index, see <http://www.lanl.gov/orgs/cic/cic6/bits/index/index_home.html>

Keywords	Title of BITS Article	Date	Page
ASCI	<i>Blue Mountain Is World's Fastest Computer</i>	Dec. 98	1
	<i>Tecolote: An Object-Oriented Framework for Hydrodynamics Physics</i>	Aug. 99	26
	<i>Using Tecolote Components to Extend Object-Oriented Programming</i>	Sept. 99	19
ASCI Training	<i>ASCI Training Course Descriptions</i>	J/J 99	27
BITS Announcements	<i>BITS Gets New Design, New Editor</i>	F/M 99	3
BITS Interviews	<i>Bob Newell</i>	O/N 98	1
	<i>Earleen Eden</i>	O/N 98	3
	<i>Dale Land</i>	Dec. 98	11
	<i>John Morrison</i>	F/M 99	6
	<i>Ann Hayes</i>	J/J 99	22
	<i>CIC Division Hosts Student Poster Presentations</i>	Sept. 99	23
<i>CIC (Computing, Information, & Communications)</i>	<i>BITS Welcomes New Deputy Division Director for Strategic Computing</i>	F/M 99	6
<i>CIC-2</i>	<i>CIC-2's Future Is in Stewardship, Automation, and "Fun Stuff"</i>	Dec. 98	11
<i>CIC-10</i>	<i>CIC-10 "Knowledge Is Our Most Important Product"</i>	O/N 98	3
Desktop Computing	<i>CIC-2's Future Is in Stewardship, Automation, and "Fun Stuff"</i>	Dec. 98	11
	<i>New Ways for Managing Software Development Projects</i>	J/J 99	2
	<i>Scientific Software Process Engineering</i>	J/J 99	4
	<i>Adjusting for Reality: Mitigating Uncertainty in Projects</i>	Aug. 99	7
	<i>Your Data Is Gone . . . Now What?</i>	Aug. 99	11
	<i>Desktop on Demand: A Desktop That Displays on Your Computer, but Actually Runs on the Enterprise Server</i>	Sept. 99	6
	<i>Grassroots Software Management: Simple Things to Do</i>	Sept. 99	8
	<i>News About Microsoft Networking at Los Alamos: The LANL Master Accounts Domain Is Up and Running!</i>	Sept. 99	10
	<i>Notes on Setting Up and Configuring an NT Domain</i>	Sept. 99	12
Embedded Systems	<i>The Year 2000 Bug Is Hiding</i>	F/M 99	1
ESD	<i>Working for and Serving You</i>	F/M 99	10
HTML	<i>Extensible HTML</i>	J/J 99	12
	<i>Web Content Architecture: Taming the Tangle of Protocols</i>	Sept. 99	15
High Performance Computing	<i>Performance and Scalability Analysis of Applications on T-flop-scale Architectures</i>	J/J 99	7
	<i>Major 3-D Parallel Simulations</i>	J/J 99	9
	<i>Lecture Review: Next-Generation Chips, Processors, Transistors, and Wiring</i>	Aug. 99	23
	<i>Tecolote: An Object-Oriented Framework for Hydrodynamics Physics</i>	Aug. 99	26

Keywords	Title of BITS Article	Date	Page
Infrastructure	<i>Virtual Private Network</i>	<i>J/J 99</i>	<i>15</i>
	<i>Electronic Information Protection Regimes</i>	<i>Aug. 99</i>	<i>13</i>
	<i>Web Content Accessibility: New W3C Guidelines Have Arrived</i>	<i>Aug. 99</i>	<i>20</i>
	<i>Web Content Architecture: Taming the Tangle of Protocols</i>	<i>Sept. 99</i>	<i>15</i>
Information Systems	<i>Engineering Index Is Now Available from the Research Library</i>	<i>Dec. 98</i>	<i>13</i>
	<i>Research Library Adds Social SciSearch® Database</i>	<i>F/M 99</i>	<i>5</i>
	<i>Library Databases</i>	<i>J/J 99</i>	<i>20</i>
	<i>aha!—An Interface to the Best of LANL Science & Technology Web Resources</i>	<i>Aug. 99</i>	<i>2</i>
	<i>More Than a New Look & Feel: The New LANL Badging System</i>	<i>Aug. 99</i>	<i>5</i>
	<i>Laboratory Overview Presentation Materials on the Web</i>	<i>Sept. 99</i>	<i>2</i>
	<i>New INSPEC® at LANL</i>	<i>Sept. 99</i>	<i>28</i>
	<i>Keep on Top of Information Technology with Gartner Group</i>	<i>Sept. 99</i>	<i>28</i>
Networks	<i>Laboratory Unclassified Network Will Implement Changes for Security</i>	<i>Dec. 98</i>	<i>3</i>
	<i>Preserving Bandwidth on Your Windows NT Network</i>	<i>F/M 99</i>	<i>8</i>
	<i>News About Networking at Los Alamos: The LANL Master Accounts Domain Is Up and Running!</i>	<i>Sept. 99</i>	<i>10</i>
	<i>Notes on Setting Up and Configuring an NT Domain</i>	<i>Sept. 99</i>	<i>12</i>
PNG	<i>It's Time for PNG: A Graphics Format You Can Pronounce</i>	<i>O/N 98</i>	<i>7</i>
Popular Science Award	<i>PS Awards LANL Web</i>	<i>O/N 98</i>	<i>12</i>
Records Management	<i>CIC-10: "Knowledge Is Our Most Important Product"</i>	<i>O/N 98</i>	<i>3</i>
SciSearch	<i>Research Library Adds Social SciSearch® Database</i>	<i>F/M 99</i>	<i>5</i>
Supercomputing	<i>Blue Mountain Is World's Fastest Computer</i>	<i>Dec. 98</i>	<i>1</i>
Virtual Private Network	<i>Virtual Private Network</i>	<i>J/J 99</i>	<i>15</i>
Wildfires	<i>Forecasting Wildfires and Other Crises</i>	<i>Apr. 98</i>	<i>3</i>
Windows NT	<i>Preserving Bandwidth on Your Windows NT Network</i>	<i>F/M 99</i>	<i>8</i>
	<i>News About Networking at Los Alamos: The LANL Master Accounts Domain is Up and Running!</i>	<i>Sept. 99</i>	<i>10</i>
	<i>Notes on Setting Up and Configuring an NT Domain</i>	<i>Sept. 99</i>	<i>12</i>
World Wide Web (WWW or Web)	<i>It's Time for PNG: A Graphics Format You Can Pronounce</i>	<i>O/N 98</i>	<i>7</i>
	<i>Lab Web Publishers Invited to IntraLab99</i>	<i>F/M 99</i>	<i>7</i>
	<i>Extensible HTML</i>	<i>J/J 99</i>	<i>1</i>
Writing Resources	<i>Telling Your Story in BITS: Sharing Your Expertise and Energy</i>	<i>Aug. 99</i>	<i>31</i>
Year 2000 (Y2K)	<i>The Year 2000 Bug Is Hiding</i>	<i>F/M 99</i>	<i>1</i>

Customer Support Center . . . (505) 665-4444, ext. 851, or cichelp@lanl.gov

Because of a wide variety of CIC computing services, numerous facilities are available to address your questions. If you are uncertain whom to call, you can always call the Customer Support Center (CSC). CSC consultants are trained to either answer your question or locate someone who can. To reach the appropriate consultant, dial 665-4444 and make your selection from the following choices:

- Option 1: New user topics including e-mail, passwords, registration, and World Wide Web
- Option 2: Enterprise Information Applications such as Travel, Time and Effort, and Purchase cards
- Option 3: Scientific computing, storage systems, and networking
- Option 4: Classroom instruction and training
- Option 5: Desktop Consulting for PC and Macintosh software and network configurations.

Consulting Via E-Mail

Customer Support	cichelp@lanl.gov
Scientific and engineering computing	consult@lanl.gov
Administrative and business computing	eiaconsult@lanl.gov
Passwords and registration	validate@lanl.gov
Macintosh computing	Mac-help@lanl.gov
PC computing	PC-help@lanl.gov
UNIX computing	UNIX-help@lanl.gov

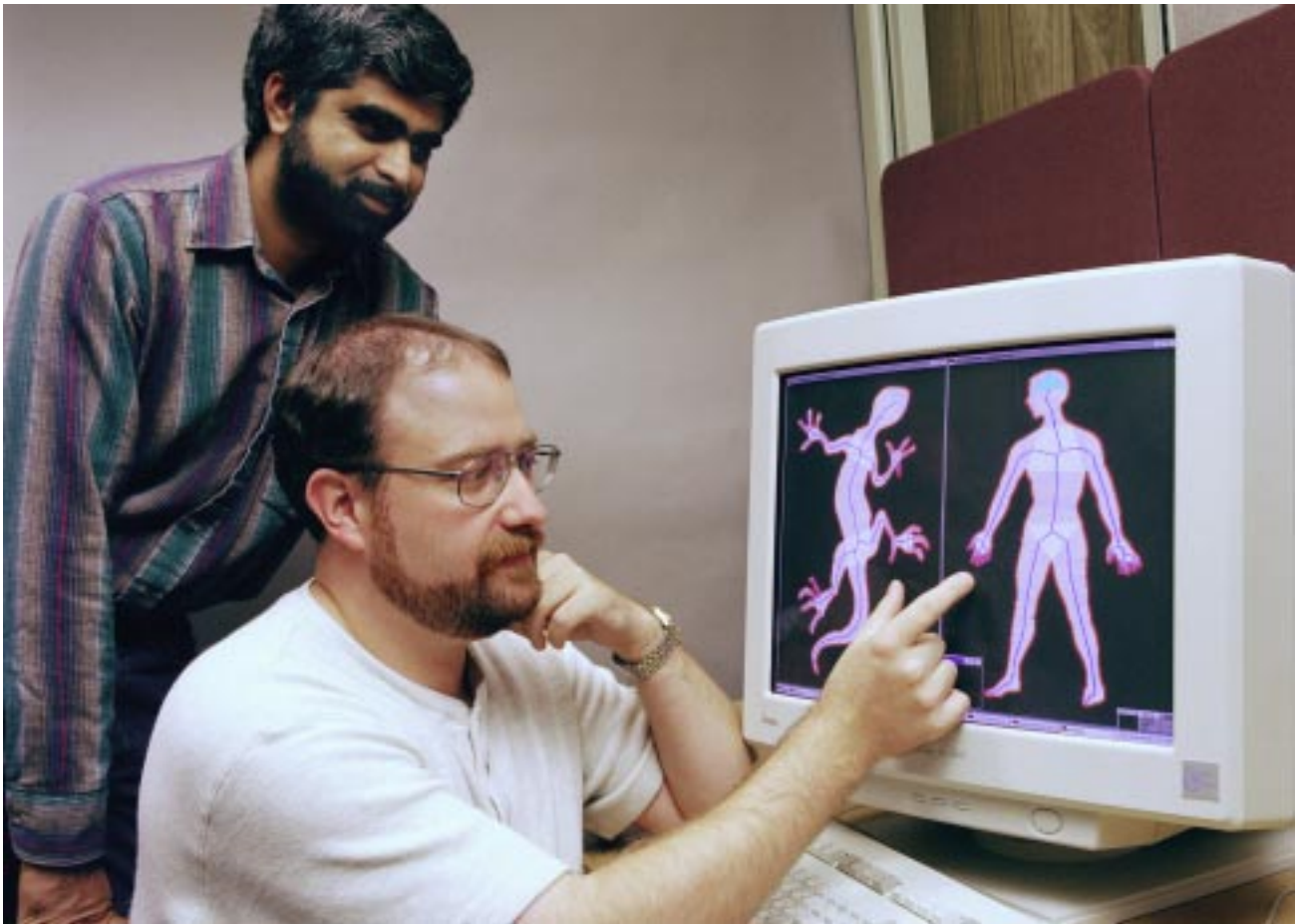
Other Useful Numbers

Advanced Computing Laboratory	665-4530
Central Computing Facility	667-4584
Network Operations Center	noc@lanl.gov or 667-7423
Telephone Services Center	667-3400

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36.

All company names, logos, and products mentioned herein are trademarks of their respective companies. Reference to any specific company or product is not to be construed as an endorsement of said company or product by the Regents of the University of California, the United States Government, the U.S. Department of Energy, nor any of their employees. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Los Alamos
NATIONAL LABORATORY



A LANL team — the Computer Vision Team — is on the forefront of research that has made, in general, great strides toward characterizing and recognizing shapes of objects in images and, in particular, distinguishing human shapes from other objects of similar shape. Lakshman Prasad, (on the left) of Nonproliferation and International Security Division's Safeguards Systems Group NIS-7, has developed a new algorithmic framework for image segmentation, object recognition, and image understanding. Coresearchers on the team are Bernd Schlei (on the right) of Theoretical Division's Equation of State and Mechanics of Materials Group T-1, and Alexei Skourikhine (not pictured) also of NIS-7.

Shape Recognition and Identification is one of the core competencies of the Advanced Surveillance Technology Initiative to identify potential threats and anomalous situations within nuclear facilities. For more information about the Advanced Surveillance Technology Initiative, contact Project Leader Sharon Seitz at (505) 665-6812 or send e-mail to sharons@lanl.gov.

Los Alamos

NATIONAL LABORATORY

Los Alamos, New Mexico 87545

A U.S. Department of Energy Laboratory

BITS is published monthly to highlight recent computing and communications activities within the Laboratory. We welcome your suggestions and contributions.

BITS may be accessed electronically at this URL:
<http://www.lanl.gov/orgs/cic/computingatlanl/>

LALP-99-42 (10/99)